Physics in Medicine & Biology

PAPER

Bag of deep features for preoperative prediction of sentinel lymph node metastasis in breast cancer

To cite this article: Jiaxiu Luo *et al* 2018 *Phys. Med. Biol.* **63** 245014

IOP Publishing Phys. Med. Biol. 63 (2018) 245014 (11pp) https://doi.org/10.1088/1361-6560/aaf241

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Bag of deep features for preoperative prediction of sentinel lymph node metastasis in breast cancer

RECEIVED
6 July 2018

REVISED
16 November 2018

ACCEPTED FOR PUBLICATION
20 November 2018

PUBLISHED
14 December 2018

Jiaxiu Luo[1], **Zhenyuan Ning**[1], **Shuixing Zhang**[2], **Qianjin Feng**[1] **and Yu Zhang**[1,3]

[1] School of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong 510515, People's Republic of China
[2] Department of Radiology, Guangdong General Hospital/Guangdong Academy of Medical Sciences, No. 106 Zhongshan Er Road, Guangzhou, Guangdong 510080, People's Republic of China
[3] Author to whom any correspondence should be addressed.

E-mail: 2529337477@qq.com, 843155331@qq.com, shui7515@126.com, qianjinfeng@gmail.com **and** yuzhang@smu.edu.cn

Keywords: convolution neural network, bag-of-features, kernel fusion, sentinel lymph node metastasis

Abstract

Breast cancer is the most common female malignancy among women. Sentinel lymph node (SLN) status is a crucial prognostic factor for breast cancer. In this paper, we propose an integrated scheme of deep learning and bag-of-features (BOF) model for preoperative prediction of SLN metastasis. Specifically, convolution neural networks (CNNs) are used to extract deep features from the three 2D representative orthogonal views of a segmented 3D volume of interest. Then, we use a BOF model to furtherly encode the all deep features, which makes features more compact and products high-dimension sparse representation. In particular, a kernel fusion method that assembles all features is proposed to build a discriminative support vector machine (SVM) classifier. The bag of deep feature model is evaluated using the diffusion-weighted magnetic resonance imaging (DWI) database of 172 patients, including 74 SLN and 98 non-SLN. The results show that the proposed method achieves area under the curve (AUC) as high as 0.852 (95% confidence interval (CI): 0.716–0.988) at test set. The results demonstrate that the proposed model can potentially provide a noninvasive approach for automatically predicting prediction of SLN metastasis in patients with breast cancer.

1. Introduction

Breast cancer is the most commonly diagnosed and prevalent female malignancy that affects women's health and remains a challenging task because of its high annual death rate (Torre *et al* 2015). Axillary lymph node (ALN) status is an important prognostic marker and provides essential guidelines of therapy decision for patients with breast cancer (Qiu *et al* 2012). Sentinel lymph node (SLN) biopsy (Krag *et al* 1993, Giuliano *et al* 1994) is introduced to reduce the morbidity of breast cancer surgery and is considered as the care standard of ALN staging, especially in negative clinical patients (Barranger *et al* 2005). Nevertheless, SLN biopsy, as an invasive operation, often leads to severe complications (Kootstra *et al* 2008). Furthermore, some histopathological and clinical data, which used as the predictors of SLN metastasis, are postoperative and cannot provide guidance for SLN biopsy (Viale *et al* 2005, Chen *et al* 2012, Ozemir *et al* 2016). Thus, accurate and noninvasive methods for the preoperative prediction of SLN status can benefit clinical surgery and therapy decision-making.

Magnetic resonance imaging (MRI) is a routine and noninvasive tool that assists clinical diagnosis and object detection for various diseases. Recent studies have shown that MRI images, including T2-weighted imaging (T2-weighted), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced imaging (DCE), are equal to detecting and predicting tasks (Li *et al* 2016, Yang *et al* 2017, Fan *et al* 2017). Functional MRI techniques, such as DWI, provide unique information about the condition of the molecular translational motion of water, and reflect different characteristics of normal tissues and tumors. However, they are rarely used for breast cancer classification or LN metastasis prediction compared with mammography, ultrasound, or anatomical MRI (Sinha *et al* 2002).

Recently, deep learning achieves outstanding success and becomes a promising approach in image processing, object detection, and many other domains (Lecun *et al* 2015). Among various deep learning methods, the

convolutional neural network (CNN) developed by Fukushima (1980) and improved by LeCun *et al* (1998) has become a hotspot and brought breakthrough in various field, such as big data and image processing. A classical CNN structure is mainly composed of three layers, namely, convolutional layer, pooling layer and full-connected layer, which are used for feature extraction, feature reduction, and classification, respectively. In contrast with traditional hand-crafted feature extraction measures, CNN learns features directly from image patches in a data-driven way. Many studies have proven that CNN is successfully used in the medical image classification and detection tasks, such as mammography breast mass (Jiao *et al* 2016), CT lung node (Ciompi *et al* 2015) and MRI prostate cancer detection (Yang *et al* 2017).

Previous studies often handle with task based on the 3D region of interest (ROI) or small 2D patches (Anthimopoulos *et al* 2016, Yang *et al* 2017). Currently, some researches have proven that decomposition view, such as sagittal, coronal and axial planes, can replace 3D to perform further analysis (Shin *et al* 2016, Vos *et al* 2016). The 2.5D approach not only can considerably reduce the computational burden, but also alleviate overfitting, especially in the medical domain (Roth *et al* 2014, 2016). The 3D CNN structure cannot train better if training data is insufficient. In this case, a 2.5D representation is generally better than 3D for specified applications.

Deep learning requires big data to train the network. However, obtaining big data for a specific medical task is difficult. Thus, some studies proposed the framework of integration of CNN-based feature descriptors and traditional classifiers such as support vector machine (SVM), which achieved great success for object detection (Huynh *et al* 2016, Wang *et al* 2017a).

Over the past decades, the bag-of-features (BOF) model has been applied in computer vision successfully. BOF involves the following three major procedures (Passali and Tefas 2017): (1) feature extraction: feature descriptors, such as SIFT (Lowe 2004), LBP (Ojala *et al* 2002), or HOG (Dalal and Triggs 2005), are extracted to encode images; (2) dictionary learning: representative features are learned as code words; and (3) object encoding: features of an image are encoded as the histogram of code words. BOF model can reduce the dimension of features significantly and select discriminative features in unsupervised way (Chatterjee *et al* 2017). Furthermore, the BOF model implements high-dimensional sparse representation, which produces more compact and typical descriptors for each image (Mohedano *et al* 2016). BOF obtains impressive success not only for natural images, but also medical image, such as detection and segmentation (Chatterjee *et al* 2017, Islam *et al* 2017, Shiji *et al* 2017).

In this paper, a DWI-based model is proposed for preoperative prediction of SLN metastasis in patients with breast cancer. The model integrates the convolution neural network and BOF model, which inspired by remarkable performance obtained by feature extractor of CNN and sparse representation of BOF scheme, respectively. Decomposition 2D orthogonal views of 3D volume of interest (VOI), namely, axial, coronal, and sagittal image planes, are used as the input of the CNN structure, and the combination of the deep features from three views is applied to represent the whole tumor. To suppress overfitting, the BOF model is used to learn codebook and encode deep features from three views, which increases the compactness and improves the representation of features. Finally, a kernel fusion strategy based on the three bags of deep features with adaptive weight is used to build a SVM classifier for classification.

2. Material and method

In the following section, the details of the proposed structure are presented. Figure 1 shows the pipeline of the proposed framework.

2.1. Dataset

In our approach, diffusion-weight images MRI are used as training and validating data. The database has 172 patients that are divided as SLN ($n = 74$) and non-SLN group ($n = 98$). Patients confirmed to breast cancer by histological diagnosis from March 2014 to June 2016 were reviewed retrospectively. SLN metastasis was also confirmed by histopathology. MRI was operated with a 1.5 T MR imager (Achieva 1.5 T, Philips Healthcare, Best, Netherlands) equipped with a four-channel SENSE breast coil in prone position. Axial DW images with bilateral breast coverage and 256 pixels $\times$ 256 pixels per slice were acquired by means of single-shot spin-echo echo-planar imaging. The other parameters are as follows: TR/TE = 5065/66 ms; FOV = 300 mm^2 $\times$ 300 mm^2; matrix = 200 $\times$ 196; slice thickness = 5 mm; slice gap = 1 mm; $\boldsymbol{b}$ value of 0 and 1000 s/mm^2.

We obtained DWI digital image and communication in medicine (DICOM) images from the picture archiving and communication system (PACS) without applying normalization. 3D VOIs were manually delineated around the whole tumors on each slice and segmented using ITK-SNAP software by experienced radiologist and inspected by a senior one. Thereafter, the window width and level were transformed into the original one, and the intensity values of the images were mapped to $[0, 1]$.

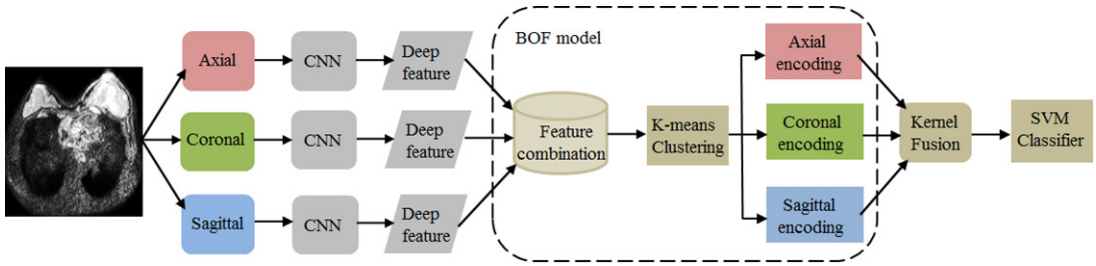

Figure 1. The pipeline of our proposed framework for SLN metastasis prediction.

2.2. CNN descriptors on representative images

2.2.1. *Represent 3D VOI by 2D orthogonal views*

CNN structure was previously trained with 2D natural images for classification or object detection. Nevertheless, different views can provide specific information in medical images. In fact, radiologists usually diagnose or provide clinical decisions according to the 2D images through the axial, coronal, and sagittal views (Ciompi *et al* 2015). In this paper, instead of using the whole 3D VOI, 2D decomposition views are addressed to represent the detection problem of SLN metastasis. The main purpose is to decrease the number of each individual input to accelerate the training process, which is less time consuming than handling with 3D images. Moreover, the integration of three representative views of the tumors is considered to use spatial information other than 2D images (Wang *et al* 2017b). Figure 2 shows that the 3D VOI is decomposed into three orthogonal views, namely, the axial, coronal, and sagittal image plane, and consider the integration of the three channels to characterize the whole tumor. In order to extract more discriminative and useful information about the disease, the slice in three views with the largest number of pixels is selected as the representative image. In this way, the process of training and testing can be accelerated and is less time consuming than 3D images. Then, to ensure the tumor information accounts for the majority of an image, the area around the tumor of each image is cropped out using a rectangular box, which making the image size to 78×78.

2.2.2. *CNN architecture*

Traditional CNN (Jiao *et al* 2016) consists of several layers with a deep supervised learning architecture that mainly includes convolutional layer, pooling layer and full-connected layer. However, in this paper, CNN is considered as a feature extractor. Thus, a simple network is used to perform on a square region, and figure 3 illustrates our extraction framework. The structure contains three convolution layers and two pooling layers similar to Tolias *et al* (2015) and Mohedano *et al* (2016), and finally, 4096 features can be extracted. The input layer is the 78×78 representative DWI slices of axial, coronal and sagittal image planes. The features of the three subsequent convolution layers are extracted from the input images or feature maps by conducting convolution process with filters. The size of filtering kernels on the feature maps is set to 3×3, and the down sampling operation on the pooling layer has a ratio of 2. The architecture becomes more complex and prompts the number of convolution layer to be increased with layer depth. Particularly, $k \times (L+1)^2$ kernels are implemented in the L_{th} convolutional layer, where k can be confirmed in the experiments. In the convolutional layer, low-level feature representations can be transferred into high-level through a non-linear activation function, such as sigmoid and rectified linear unit (ReLU), and the type of the activation is also determined by experiments. Between convolution layers, pooling operation is performed to summarize the feature responses across adjacent pixels (Roth *et al* 2016) and the crucial information, which realizes the reduction of feature numbers and achieves translation, rotation, and scale invariance, is selected. The number of pooling layer is equal to the previous convolutional layer because the subsampling function only changes the size of the feature maps. The type of pooling and other parameters of the CNN are defined through the experiments.

2.3. Generation of the bag of deep features

2.3.1. *Feature concatenation*

Our approach works by creating features on three orthogonal views through CNN structure. The integration of axial, coronal, and sagittal representational information, which contain different spatial information, can characterize the whole tumor, and the combination of 2D views can provide insights on the type of object detected in a 3D image. Hence, after feature extraction, the three descriptors are concatenated to form the feature set as

$$f_{combined} = \{f_a, f_c, f_s\} \tag{1}$$

where f_a, f_c, f_s are the three descriptors that correspond to axial, coronal, and sagittal views, respectively.

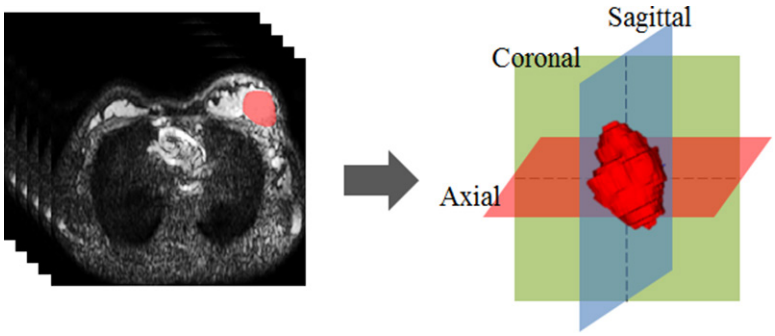

Figure 2. Example of a segmented 3D VOI and decomposition with three orthogonal views. The slice with largest region of axial, coronal and sagittal slices in each patient is chosen to represent the three views, respectively.

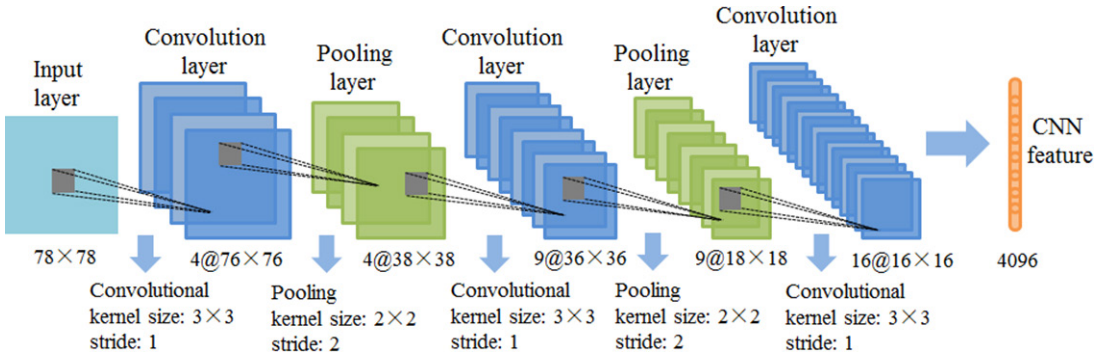

Figure 3. Architecture of CNN used for feature extraction.

2.3.2. Bag of deep features

High-dimension features are always time consuming and result in overfitting for classification task. Consequently, acquiring the most remarkable features by reducing the dimension of features is necessary. In our work, we employ the BOF model, which can select relevant features and encode the images with sparse representation, to reduce feature dimension. Furthermore, BOF-based features are more compact and easy to interpret (Mohedano *et al* 2016).

Figure 4 shows the flowchart of our model. In the training step, the codebook is built on the basis of the training data using *k*-means cluster algorithm, and *k* words are the centroids of each cluster. *K*-means, which quantizes the features and forms a codebook, is widely applied in image processing. Therefore, each representation of a CNN feature is assigned to its closest codeword within the codebook, and images are encoded on the basis of the frequency of the codewords, where the counts are represented by a histogram. This process reduces high-dimensional features related to the original image to only *k* representative features.

Thus, images are sparse represented with bag of deep features. In the testing process, feature extraction is the same as the training set and directly uses the codebook acquired in the training process to encode the testing images. Finally, the testing features are conducted into classification equally to the training set.

2.4. SVM classifier based on kernel combination

After executing the BOF algorithm, each image is represented as a vector with k-dimension. The integration of axial, coronal, and sagittal views of the 3D VOI is regarded as the representation of the whole tumor. Thus, our classification is based on the combination of the three decompositions. SVM has become popular due to its significant empirical performance in classification works in recent decades. The binary category of SLN prediction in our work is classified using a SVM, and all parameters of the SVM classifier are obtained through parallel training. For most data, the use of non-linear SVM model, which has kernel function to transform features into a high-dimensional space to express the data more precise and figurative, is preferred. Moreover, many researchers deal with object detection using conventional classification, in which SVM has only one kernel and the capability is limited (Huynh *et al* 2016, Hou *et al* 2016).

However, the three feature sets cannot be used fully to build a discriminative SVM because all the features are simply combined as inputs. Features from three orthogonal views may contain different spatial information and account for different proportions when the whole tumor is characterized. In our approach, we use kernel fusion method (Tao *et al* 2016) to combine the three representative features (figure 4) to improve the traditional kernel

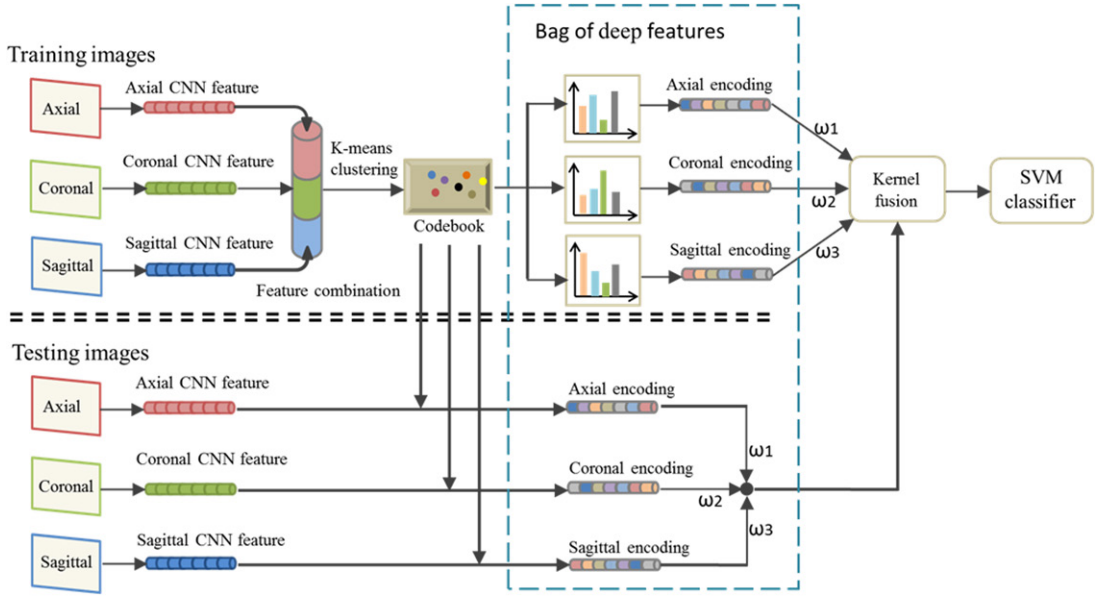

Figure 4. The flowchart of the BOF model to encode the axial, coronal and sagittal images and kernel fusion is conducted to train a discriminative SVM.

calculating method. In conventional SVM classification, a feature vector from 3D images is used to compute kernel function, but in the present work, we conduct three decomposition views as 3D representation. Thus, three representative features are combined with different coefficients into a kernel computation to express the tumor accurately. This method indicates that the three feature channels are assembled with the mixing coefficients as $\omega = \{\omega_1, \omega_2, \omega_3 | \omega_1 + \omega_2 + \omega_3 = 1\}$, where subscripts 1, 2, and 3 denotes the axial, coronal and sagittal coefficients, respectively. SVM is trained in tenfold cross-validation and the three ratios can be identified according to experiments on the validation set. Specially, different features usually fit to different kernels due to the diverse structure of the data, and the choice of the kernel type is crucial and depends on the data distribution and specific application. Therefore, the use of several frequently used kernel types (see table 1) is proposed to deal with a nonlinear classification and to compare these kernel types based on their performance.

The decision function in testing images for different kernel functions is defined as follows:

$$f(x) = sign\left(\sum_{i=1}^{M} \alpha \left(\sum_{v=1}^{3} \omega_v K_v(x_i, x)\right) + b\right) \tag{2}$$

where K_v means the kernels of different image planes; $x_i \in [I_0, I_1]^T$ and I_0, I_1 are the feature vector of the training images in negative and positive labels, respectively; x represents the testing images; α is the weight parameter, and b denotes the learned intercepts determined by training experiments; and $f(x)$ is the final prediction label of image x; M is the number of training images.

3. Experiment setup and result

3.1. Experiment setup

The experiment approach is implemented on three datasets, including training set ($n = 100$), validation set ($n = 22$), and testing set ($n = 50$). The training set is used to train the whole structure, and parameter fine-tuning is based on the validation set. The testing set is applied to access the effectiveness of our model. A linear normalization method is used to eliminate the magnitude of features and negative effects of large magnitude difference in the following definition:

$$F_{norm} = \frac{F - F_{\min}}{F_{\max} - F_{\min}} \tag{3}$$

where F denotes the features. To estimate the prediction performance, AUC, which is the area under the receiver operating characteristic (ROC), is used as the representation. ROC curve uses true-positive rate (TPR) and false-positive rate (FPR) as the ordinate and abscissa axes respectively, which are defined as:

$$TPR = \frac{TP}{TP + FN}; FPR = \frac{FP}{TN + FP} \tag{4}$$

where TP, TN, FP, and FN correspond to true positive, true negative, false positive and false negative.

Table 1. Different kernel functions for kernel combination.

Kernel	Kernel fusion function	Parameters
Linear	$K_{\mathrm{LIN}}(X_i, X_j) = \sum_{v=1}^{3} \omega_v F_{\mathrm{LIN}}^{v}(X_i, X_j)$	$F_{\mathrm{LIN}}(X_i, X_j) = \sum_{k=1}^{N} (X_i^k)^T \times X_j^k$
Polynomial	$K_{\mathrm{POL}}(X_i, X_j) = (1 + \sum_{v=1}^{3} \omega_v F_{\mathrm{POL}}^{v}(X_i, X_j))^d$	$F_{\mathrm{POL}}(X_i, X_j) = \sum_{k=1}^{N} X_i^k \times X_j^k$
Gaussian (RBF)	$K_{\mathrm{RBF}}(X_i, X_j) = \exp\left(\frac{-\sum_{v=1}^{3} \omega_v D_{\mathrm{RBF}}^{v}(X_i, X_j)}{2\sigma^2}\right)$	$D_{\mathrm{RBF}}(X_i, X_j) = \sum_{k=1}^{N} (X_i^k - X_j^k)^2$
Sigmoid	$K_{\mathrm{SIG}}(X_i, X_j) = \tanh(\beta \times \sum_{v=1}^{3} \omega_v F_{\mathrm{SIG}}^{v}(X_i, X_j) - \theta)$	$F_{\mathrm{SIG}}(X_i, X_j) = \sum_{k=1}^{N} X_i^k \times X_j^k$
Intersection (HIK)	$K_{\mathrm{HIK}}(X_i - X_j) = \sum_{v=1}^{3} \omega_v F_{\mathrm{HIK}}^{v}(X_i, X_j)$	$F_{HIK}(X_i, X_j) = \sum_{k=1}^{N} \min(X_i^k - X_j^k)$
Chi-square (χ^2)	$K_{\chi^2}(X_i, X_j) = \exp\left(\frac{-\sum_{v=1}^{3} \omega_v D_{\chi^2}^{v}(X_i, X_j)}{2\sigma^2}\right)$	$D_{\chi^2}(X_i, X_j) = \frac{1}{2}\sum_{k=1}^{N} \frac{(X_i^k - X_j^k)^2}{\lvert X_i^k + X_j^k \rvert}$

Note. $v = 1, 2, 3$ responds to axial, coronal and sagittal representation; X_i and X_j are the representation of the ith and jth training data; σ, d, β, θ are hyper-parameters determined on experiments; D is the distance function between two objects; F is the basic kernel function between two objects; where k represents the kth feature in the feature vector X_i and X_j; N is the dimension of image features.

3.2. Results

3.2.1. *Tuning of CNN architecture*

The CNN architecture is used as a feature extractor in our approach, and parameter setting is important for the whole framework. The effect of different parameters for the configuration is demonstrated, and the assembly with the best performance is selected to form the final CNN structure. Table 2 shows the classification performance of different parameter combinations, and the proposed network that yields the best validation AUC of 0.883 is presented in bold. As describe in the CNN architecture section, k multiplier is used to determine the number of kernels. To identify the optimal number, we experiment with several values, and the results demonstrate that 1 is the best choice. When resolving the depth of the CNN structure, we regard the convolution layer and pooling layer as independent, and the final structure with three convolution layers and two pooling layers (see figure 3) achieves the significant classification result. Moreover, the increase in kernel size from 3×3 to 7×7 results in performance drop, which is mainly due to the discriminative texture feature that can be captured with small receptive field. Examples of kernels with different sizes and the last convolution layer images are shown in figure 5. Large kernel sizes lead to fuzzy margin and blurry structure. Two types of activation function and pooling methods are used in the experiments, and ultimately ReLU activation and max pooling type are applied in consideration of the validation of AUC.

3.2.2. *Determination of the codebook size in BOF model*

We use the aforementioned BOF model to reduce the number of CNN features and encode the images with the codebook, which is regarded as sparse representation. In the BOF framework, the feature selection process is determined by k-means clustering, and the number of vocabularies is the substantial factor that affects classification accuracy. Thus, analyzing how many centroids fit is important to improve the performance on our images. We compare the performance in our work with various dimensions of codebook from 1 to 28 assessed by the AUC of validation, whereas a codebook of large dimension causes overfitting and is more likely to indicate codeword redundancy. The comparison results are illustrated in figure 6. As we can see from the figure, size of 8 achieves the best result among the 28 clusters (AUC = 0.883). Thus, in our work, all features are clustered to 8 centroids, and the three image planes are encoded into a feature vector with dimension of 8.

3.2.3. *Relationship between mixing coefficient and kernel fusion method*

Instead of arranging the features ordinarily to train a SVM classifier to make prediction, we combine the three represented descriptors with adaptive weight into one kernel to build a discriminative SVM. These three features contain different characters that have different roles in describing the whole VOI. Classification accuracy is related with the mixing coefficients ($\omega 1, \omega 2, \omega 3$) and the selected kernel function. To identify the influence of the three coefficients on the final decision with different kernels, we set values for $\omega 1$ and $\omega 2$ from 0 to 1 in an interval of 0.05. $\omega 3$ is ensured by $\omega_1 + \omega_2 + \omega_3 = 1$ (Yuan *et al* 2015).

Figure 7 shows the relationship between the coefficients with different kernel types. For the significant AUC result, the axial-view feature holds the largest proportion among the three coefficients, and coronal take the second place in all kernel types. In our experimental data, the axial image plane has larger ROI than the coronal and sagittal planes. Subsequently, more discriminative information can be extracted from the axial view, which has

Table 2. Performance of CNN architecture for different parameters.

Number of kernels	Number of layers	Kernel size	Activation	Pooling type	Validation AUC
$k = 1$	**5**	**3**	**Relu**	**Max**	**0.883**
$k = 2$	5	3	Relu	Max	0.843
$k = 3$	5	3	Relu	Max	0.877
$k = 4$	5	3	Relu	Max	0.863
$k = 5$	5	3	Relu	Max	0.832
$k = 1$	2	3	Relu	Max	0.852
$k = 1$	3	3	Relu	Max	0.868
$k = 1$	4	3	Relu	Max	0.877
$k = 1$	**5**	**3**	**Relu**	**Max**	**0.883**
$k = 1$	6	3	Relu	Max	0.868
$k = 1$	**5**	**3**	**Relu**	**Max**	**0.883**
$k = 1$	5	4	Relu	Max	0.863
$k = 1$	5	5	Relu	Max	0.852
$k = 1$	5	6	Relu	Max	0.868
$k = 1$	5	7	Relu	Max	0.843
$k = 1$	**5**	**3**	**Relu**	**Max**	**0.883**
$k = 1$	5	3	Sigmoid	Max	0.868
$k = 1$	**5**	**3**	**Relu**	**Max**	**0.883**
$k = 1$	5	3	Relu	Mean	0.863

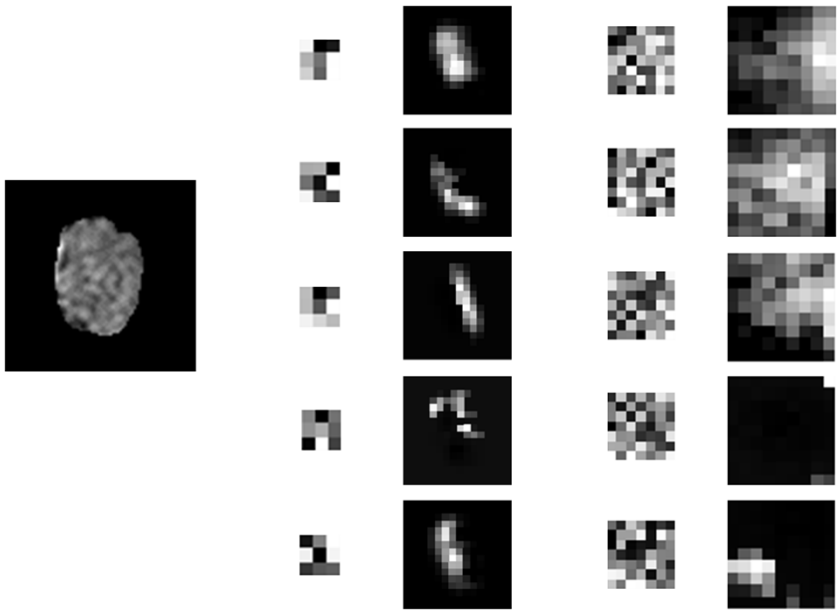

Figure 5. Examples of the last convolutional kernels with different size (2th column: 3×3 and 4th column: 7×7) and responses of the last convolution layer (3th column and 5th column); the image in the first column is an axial example.

the largest contribution. Another reason for this distribution is that the resolution of the axial views is better than the other two planes.

Table 3 provides the best coefficient integration in six kernel types with the highest validation AUC. The chi-square kernel with the coefficients of 0.55, 0.3, and 0.15 obtains the best outcome of 0.883. As described previously, the feature extraction process is conducted on the ROIs, and therefore, our features are all non-negative. Moreover, we normalize the feature vectors to the range [0, 1], therefore, the feature structure is similar to the probability distribution function (PDF) and histogram-like characters (Adeli *et al* 2017). Chi-square kernel function, which computes the χ^2 distance between two objects, is a distance metric used for this type of architecture (Cabral *et al* 2015). Consequently, chi-square kernel function is more suitable for our data and the result demonstrates the conclusion. Intersection kernel, which capture the similarities of objects can also acquire an outstanding outcome, is also a popular kernel used for histogram-like features (Adeli *et al* 2017). However,

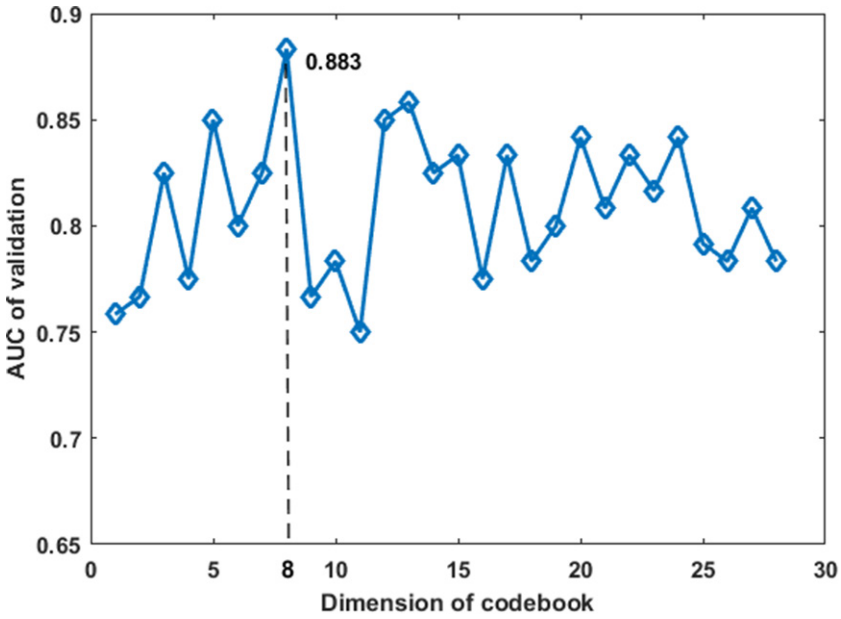

Figure 6. AUC of the validation based on different dimensions of codebook from 1 to 28 and the number 8 achieves the best performance with AUC of 0.883.

Table 3. The best performance with the mixing coefficients in each kernel type.

Kernel	Coefficients	Validation AUC
Linear	$\omega_1 = 0.70, \omega_2 = 0.20, \omega_3 = 0.10$	0.850
Polynomial	$\omega_1 = 0.70, \omega_2 = 0.25, \omega_3 = 0.05$	0.842
Gaussian	$\omega_1 = 0.60, \omega_2 = 0.25, \omega_3 = 0.15$	0.867
Sigmoid	$\omega_1 = 0.55, \omega_2 = 0.35, \omega_3 = 0.10$	0.833
Intersection	$\omega_1 = 0.80, \omega_2 = 0.10, \omega_3 = 0.10$	0.875
Chi-square	$\boldsymbol{\omega_1 = 0.55, \omega_2 = 0.30, \omega_3 = 0.15}$	**0.883**

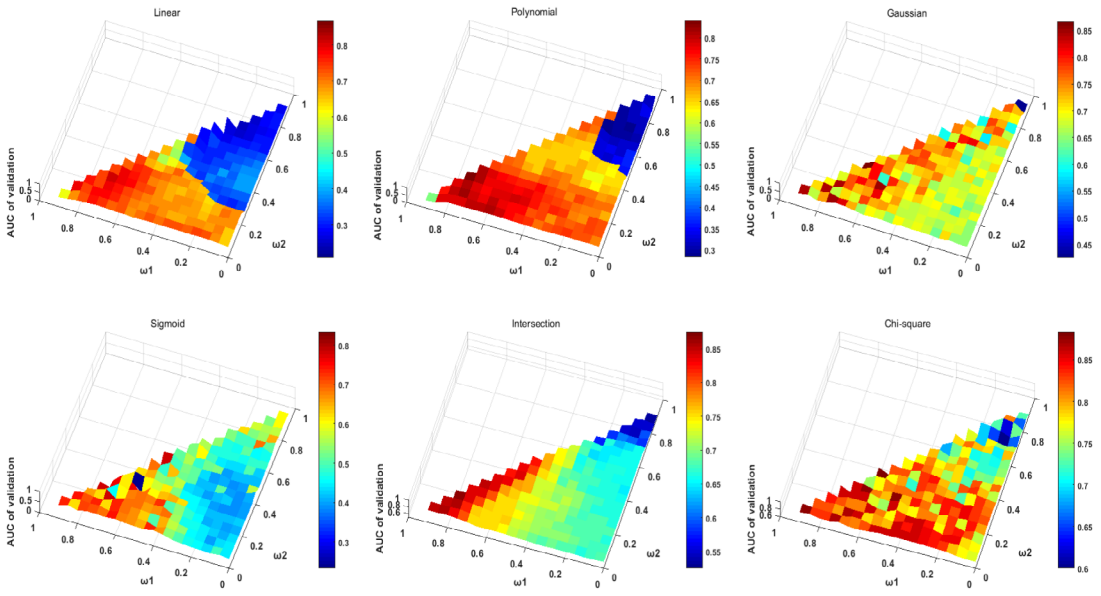

Figure 7. Relationship between the three coefficients and performance with six kernel types.

Table 4. Performance of the three views and combined feature with and without BOF model.

Method	Testing AUC	Accuracy	Sensitivity	Specificity
Axial + SVM	0.798(0.661–0.898)	0.780	0.600	0.800
Axial + BOF + SVM	0.828(0.695–0.920)	0.820	0.800	0.833
Coronal + SVM	0.767(0.626–0.875)	0.740	0.600	0.733
Coronal + BOF + SVM	0.788(0.650–0.891)	0.760	0.650	0.800
Sagittal + SVM	0.737(0.593–0.851)	0.720	0.550	0.733
Sagittal + BOF + SVM	0.777(0.637–0.882)	0.740	0.600	0.767
Combined feature + SVM	0.823(0.689–0.917)	0.800	0.750	0.867
Combined feature + BOF + SVM	**0.852(0.716–0.988)**	**0.840**	**0.867**	**0.833**

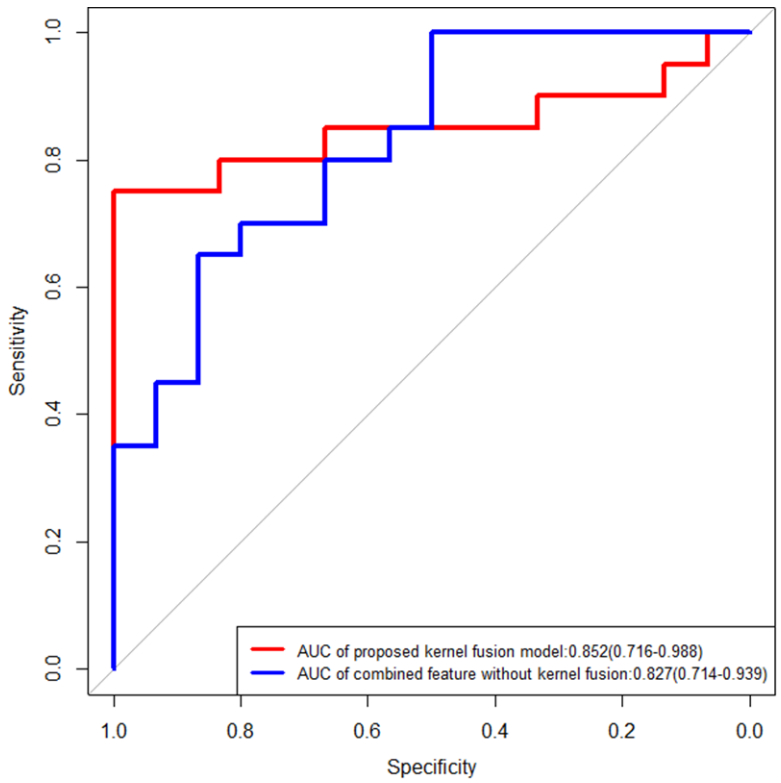

Figure 8. Performance of feature combination without kernel fusion and the proposed kernel fusion model evaluated by AUC of test set.

chi-square is the best choice in our work. When the model in the testing set is finally implemented, we can yield an AUC of 0.852, showing that the framework has prominent capability for SLN status prediction.

4. Discussion

4.1. Effectiveness of BOF model and kernel fusion method

In this paper, we develop the aforementioned BOF structure to achieve a great performance. To identify the effectiveness of BOF model, we compare the performances of the proposed method with three single orthogonal views and the combined features without the BOF model. Table 4 shows the results of our experiments. As expected, the combined features obtain a higher testing AUC than the three single 2D orthogonal views. The axial images, also shows the best performance among the three views. Furthermore, the result is improved by using the BOF model. The best performance can be obtained when using kernel fusion strategy, which demonstrates the effect of the proposed method.

In figure 8, we explore the effect of kernel fusion. Feature combination without kernel fusion is inferior to the performance of the proposed fusion model. The result indicates that the integration of information from the three orthogonal views is an impactful strategy. In summary, the evaluated bag of deep feature model and kernel fusion method can provide significant testing results in SLN prediction.

Table 5. Comparisons of different methods.

Method	Testing AUC
3D + BOF + SVM	0.815
Dong *et al*	0.787
Proposed method	**0.852**

4.2. Comparisons with other methods

Conventional CNN often needs big data to train and it is difficult for a specific medical task. Therefore, many studies utilize patch-based strategy to deal with such issues (Anthimopoulos *et al* 2016, Cordier *et al* 2016). Nonetheless, patch-based approaches may cause confusion in clinical decision and features are only from a local region. The ignored global features, such as volume and size of tumor, are also significant in clinical practice. Many literatures have proposed CNN-based descriptors fed into traditional classifiers to remit the requirement of big data (Huynh *et al* 2016, Mohedano *et al* 2016).

We compare our proposed method with 3D-based method. As shown in table 5, 3D-based method obtains AUC of 0.815, and the proposed method gets better AUC of 0.852. For 3D-based method, insufficient training data may cause scalability issue and often cannot train model effectively. Using three decomposition views and traditional classifier can remit the requirement of big data and accelerate the training process. Moreover, the integration of three representative views of the tumors can make the best of spatial information and the result validates that it generalizes better than 3D-based method.

We also compare the proposed method with the previous method for the same task. In Dong *et al* (2017), radiomics method based on texture and non-texture features are identified for the preoperative prediction of SLN metastasis. The model with radiomics features extracted from DWI images reached an AUC of 0.847 in the training set and 0.787 in the validation set. For joint anatomical and functional MRI images (T_2-FS and DWI), the optimal model just yielded an AUC of 0.805 in the validation set. Consequently, the proposed method can achieve better performance and potentially provide a non-invasive approach in clinical practice.

5. Conclusion

A novel structure that integrates CNN and BOF model is proposed in this paper for prediction of SLN metastasis. This structure also applies kernel fusion to train a discriminative SVM classifier. Our method uses three decomposition orthogonal views (axial, coronal and sagittal views) to replace 3D VOI. The CNN structure acts as a feature descriptor and is used to extract features from the three representative image planes. After obtaining the CNN features, the direct concatenation of the deep features is fed to the BOF model. Then, we use k-means clustering algorithm based on BOF to form a codebook, which increases the compactness of features, and produce high-dimensional sparse representation. For classification, three descriptors are fused with adaptive weight to train SVM classifier. The experimental results demonstrate that our method achieves promising success in SLN status prediction. This framework synthesizes two robust technologies, i.e. CNN for feature extraction and BOF model for features selection. Nevertheless, our approach works based on manual tumor segmentation, which is time consuming. Automatic image segmentation needs further study. Meantime, during the training process of CNN, the three CNN structures are trained for feature extraction from axial, coronal and sagittal images individually. We just used cross-entropy as the loss function and there is no comparison with other loss functions. In the future, we will further study CNNs based on different loss function and optimization algorithm. And we are dedicated to improving the performance of our model by fine-tuning the architecture and parameters, which can provide a more accuracy noninvasive model for predicting SLN metastasis in patients with breast cancer.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant No. 61671230 and No. 31271067, the Science and Technology Program of Guangdong Province under Grant No. 2017A020211012, the Guangdong Provincial Key Laboratory of Medical Image Processing under Grant No. 2014B030301042, and the Science and Technology Program of Guangzhou under Grant No. 201607010097.

References

Adeli E, Wu G, Saghafi B, An L, Shi F and Shen D 2017 Kernel-based joint feature selection and max-margin classification for early diagnosis of parkinson's disease *Sci. Rep.* **7** 41069

Anthimopoulos M, Christodoulidis S, Ebner L, Christe A and Mougiakakou S 2016 Lung pattern classification for interstitial lung diseases using a deep convolutional neural network *IEEE Trans. Med. Imaging* **35** 1207–16

Barranger E, Coutant C, Flahault A, Delpech Y, Darai E and Uzan S 2005 An axilla scoring system to predict non-sentinel lymph node status in breast cancer patients with sentinel lymph node involvement *Breast Cancer Res. Treat.* **91** 113–9
Cabral R, Torre F D L, Costeira J P and Bernardino A 2015 Matrix completion for weakly-supervised multi-label image classification *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 121–35
Chatterjee S, Dey N, Shi F, Ashour A S, Fong S J and Sen S 2017 Clinical application of modified bag-of-features coupled with hybrid neural-based classifier in dengue fever classification using gene expression data *Med. Biol. Eng. Comput.* **56** 709–20
Chen J Y *et al* 2012 Predicting sentinel lymph node metastasis in a Chinese breast cancer population: assessment of an existing nomogram and a new predictive nomogram *Breast Cancer Res. Treat.* **135** 839–48
Ciompi F, Hoop B D, Riel S J V, Chung K, Scholten E T, Oudkerk M, Jong P A D, Prokop M and Ginneken B V 2015 Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box *Med. Image Anal.* **26** 195–202
Cordier N, Delingette H and Ayache N 2016 A patch-based approach for the segmentation of pathologies: application to glioma labelling *IEEE Trans. Med. Imaging* **35** 1066–76
Dalal N and Triggs B 2005 Histograms of oriented gradients for human detection *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition* vol 1 pp 886–93
Dong Y *et al* 2017 Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of t2-weighted fat-suppression and diffusion-weighted mri *Eur. Radiol.* **28** 582–91
Fan M, Wu G, Cheng H, Zhang J, Shao G and Li L 2017 Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients *Eur. J. Radiol.* **94** 140–7
Fukushima K 1980 Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position *Biol. Cybern.* **36** 193–202
Giuliano A E, Kirgan D M, Guenther J M and Morton D L 1994 Lymphatic mapping and sentinel lymphadenectomy for breast cancer *Ann. Surg.* **220** 391–401
Hou L, Samaras D, Kurc T M, Gao Y, Davis J E and Saltz J H 2016 Patch-based convolutional neural network for whole slide tissue image classification *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition* pp 2424–33
Huynh B Q, Li H and Giger M L 2016 Digital mammographic tumor classification using transfer learning from deep convolutional neural networks *J. Med. Imaging* **3** 034501
Islam M, Dinh A V and Wahid K A 2017 Automated diabetic retinopathy detection using bag of words approach *J. Biomed. Sci. Eng.* **10** 86–96
Jiao Z, Gao X, Wang Y and Li J 2016 A deep feature based framework for breast masses classification *Neurocomputing* **197** 221–31
Kootstra J, Hoekstra-Weebers J E H M, Rietman H, Vries J D, Baas P, Geertzen J H B and Hoekstra H J 2008 Quality of life after sentinel lymph node biopsy or axillary lymph node dissection in stage I/II breast cancer patients: a prospective longitudinal study *Ann. Surg. Oncol.* **15** 2533–41
Krag D N, Weaver D L, Alex J C and Fairbank J T 1993 Surgical resection and radiolocalization of the sentinel lymph node in breast cancer using a gamma probe *Surg. Oncol.* **2** 335–40
Lecun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436
Lecun Y, Bottou L, Bengio Y and Haffner P 1998 Gradient-based learning applied to document recognition *Proc. IEEE* **86** 2278–324
Li H *et al* 2016 MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of mammaprint, oncotype DX, and PAM50 gene assays *Radiology* **281** 382–91
Lowe D G 2004 Distinctive image features from scale-invariant keypoints *Int. J. Comput. Vis.* **60** 91–110
Mohedano E, Salvador A, McGuinness K, Marques F, O'Connor N E and Giro-i Nieto X 2016 Bags of local convolutional features for scalable instance search *Proc. ACM Int. Conf. on Multimedia Retrieval (ACM) (New York, 6–9 June 2016)* pp 327–31
Ojala T, Pietikainen M and Maenpaa T 2002 Multiresolution gray-scale and rotation invariant texture classification with local binary patterns *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 971–87
Ozemir I A, Orhun K, Eren T, Baysal H, Sagiroglu J, Leblebici M, Ceyran A B and Alimoglu O 2016 Factors affecting sentinel lymph node metastasis in Turkish breast cancer patients: predictive value of Ki-67 and the size of lymph node *Bratisl. Lek. Listy* **117** 436–41
Passali N and Tefas A 2017 Neural bag-of-features learning *Pattern Recognit.* **64** 277–94
Qiu P F, Liu J J, Wang Y S, Yang G R, Liu Y B, Sun X, Wang C J and Zhang Z P 2012 Risk factors for sentinel lymph node metastasis and validation study of the MSKCC nomogram in breast cancer patients *Japan. J. Clin. Oncol.* **42** 1002–7
Roth H R, Lu L, Liu J, Yao J, Seff A, Cherry K, Kim L and Summers R M 2016 Improving computer-aided detection using convolutional neural networks and random view aggregation *IEEE Trans. Med. Imaging* **35** 1170–81
Roth H R, Lu L, Seff A, Cherry K M, Hoffman J, Wang S, Liu J, Turkbey E and Summers R M 2014 A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations *Med. Image Comput. Comput.-Assist. Intervention* **17** 520–7
Shiji T P, Remya S and Thomas V 2017 Computer aided segmentation of breast ultrasound images using scale invariant feature transform (SIFT) and bag of features *Proc. Comput. Sci.* **115** 518–25
Shin H C, Roth H R, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D and Summers R M 2016 Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning *IEEE Trans. Med. Imaging* **35** 1285–98
Sinha S, Lucas-Quesada F A, Sinha U, DeBruhl N and Bassett L W 2002 *In vivo* diffusion-weighted MRI of the breast: potential for lesion characterization *J. Magn. Reson. Imaging* **15** 693–704
Tao Q Q, Zhan S, Li X H and Kurihara T 2016 Robust face detection using local CNN and SVM based on kernel combination *Neurocomputing* **211** 98–105
Tolias G, Sicre R and Jégou H 2015 Particular object retrieval with integral max-pooling of CNN activations *Proc. Int. Conf. Learning Representations (ICLR)* (arXiv.1511.05879)
Torre L A, Bray F, Siegel R L, Ferlay J, Lortet-Tieulent J and Jemal A 2015 Global cancer statistics, 2012 *CA-A Cancer J. Clin.* **65** 87–108
Viale G, Zurrida S, Maiorano E, Mazzarol G, Pruneri G, Paganelli G, Maisonneuve P and Veronesi U 2005 Predicting the status of axillary sentinel lymph nodes in 4351 patients with invasive breast carcinoma treated in a single institution *Cancer* **103** 492–500
Vos B D D, Wolterink J M, Jong P A D, Viergever M A and Išgum I 2016 2D image classification for 3D anatomy localization: employing deep convolutional neural networks *Medical Imaging 2016: Image Processing Int. Society for Optics and Photonics* **9784** 97841Y
Wang J, Huang P, Huang Q, Ke Z and Lin P 2017a Dialogue act recognition for Chinese out-of-domain utterances using hybrid CNN-RF *Int. Conf. on Asian Language Processing IEEE* pp 14–7
Wang X, Guo Y, Wang Y and Yu J 2017b Automatic breast tumor detection in abvs images based on convolutional neural network and superpixel patterns *Neural Comput. Appl.* **1** 1–13
Yang X, Liu C, Wang Z, Yang J, Min H L, Wang L and Cheng K T 2017 Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI *Med. Image Anal.* **42** 212–27
Yuan L, Chen F, Zhou L and Hu D 2015 Improve scene classification by using feature and kernel combination *Neurocomputing* **170** 213–20